'Listening' to Dyslexic Children's Reading: The Transcription and Segmentation Accuracy for ASR

Husniza Husni[1], Nik Nurhidayat Nik Him[1], Mohamed M. Radi[2], Yuhanis Yusof[1], Siti Sakira Kamaruddin[1]
[1]*Human-Centered Computing Research Lab, School of Computing, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*
[2]*Emirates Canadian University College, Umm Al Quwain, United Arab Emirates*
husniza@uum.edu.my

***Abstract*—Dyslexic children read with a lot of highly phonetically similar error that is a challenge for speech recognition (ASR). Listening to the highly phonetically similar errors are indeed difficult even for a human. To enable a computer to 'listen' to dyslexic children's reading is even more challenging as we have to 'teach' the computers to recognize the readings as well as to adapt to the highly phonetically similar errors they make when reading. This is even more difficult when segmenting and labelling the read speech for processing prior to training an ASR. Hence, this paper presents and discusses the effects of highly phonetically similar errors on automatic transcription and segmentation accuracy and how it is somehow influenced by the spoken pronunciations. A number of 585 files of dyslexic children's reading is used for manual transcription, force alignment, and training. The recognition of ASR engine using automatic transcription and phonetic labelling obtained an optimum result, which is with 23.9% WER and 18.1% FAR. The results are almost similar with ASR engine using manual transcription 23.7% WER and 17.9% FAR.**

***Index Terms*—Automatic Transcription and Phonetic Labelling; Automatic Speech Recognition; Dyslexic Children Reading.**

I. INTRODUCTION

Automatic speech recognition (ASR) can play an important role in boosting children's interest in learning to read using computers. The availability of ASR technology gives the opportunity to help children especially dyslexics to enhance their learning ability by using Automatic Reading Tutor (ART) or Interactive Reading Tutor (IRT). This work is a revisit work to the existing technologies and techniques, but it aims to focus more on dyslexic children read a speech with highly phonetically similar errors, which remains a challenge for ASR to accurately recognize sounds phonetically. However, this work concerns more on the investigation of whether or not automatic transcription and segmentation could produce somewhat similar accuracy to the manual counterpart when they are used as input for training an ASR. It is important to enable automatic transcription and segmentation, as in manual it would be too cumbersome to handle, especially when dealing with larger corpus with more phonetically similar pronunciations.

In order to develop ART or IRT using ASR technology, speech samples of dyslexic children's reading aloud are used to perform transcription and phonetic labelling that serve as important basic elements for the construction of an ASR engine [1- 6].

Since transcription and phonetic labelling are used to develop ASR engines, the training and its accuracy evaluation must be done by using standard methods and metrics (e.g. hybrid Hidden Markov Model (HMM) and Artificial Neural Network (ANN) for training; Word Error Rate (WER) and False Alarm Rate (FAR) for measuring accuracy). However, in this study, the dyslexic children's read speech presents a challenge to perform accurate transcription and phonetic labelling due to highly phonetically similar errors that affected the accuracy of an ASR engine. Some of the highly phonetically similar errors are presented in Table 1. These errors are made when dyslexic children were reading aloud the words and their readings were recorded and later transcribed. The errors are highly phonetically similar especially when dealing with vowel substitutions, consonant substitutions, and nasal removal, as a few examples. This situation creates a challenge for ASR and even automatic transcription and phonetic labelling to obtain acceptable accuracy, which is important in any ASR application and for automatically performing the transcription and phonetic labelling prior to training an ASR engine.

Table 1
Sample of Highly Phonetically Similar Reading Mistakes

Original word	Sample of phonetically similar error	Error type
kemarau	kemarai	Vowel substitution
	kemaru	Vowel deletion
	kemurai	Vowel substitution
cendawan	sendawan	Consonant substitution
	cedawan	Nasal removal (remove n)
	dedawan	Consonant substitution, Nasal removal
maklumat	makulmat	Incorrect sequence (u and l)
	mak umat	Liquids removal (remove l)
	mak long	Word substitution
binatang	bintang	Vowel deletion
	natang	Syllable deletion
	pinatang	Consonant substitution
abang	adang	Letter reversal
	abing	Vowel substitution
	adangan	Letter reversal, syllable addition

The investigation of performance accuracy starts with producing transcription and phonetic labelling by both manually and automatically. Based on previous studies, researchers believed that the level of accuracy when using manual transcription and phonetic labelling is higher [7-12]. The reason behind this is because the procedure of manual transcription requires human transcribers to hear the sound of each phoneme before performing transcription and phonetic labelling thus contributing to a more accurate task when compared with automatic transcription and phonetic

labelling. Even though manual transcription has shown a remarkable accuracy of spoken utterances, the accuracy performance of automatic transcription and phonetic labelling still need to be examined. This is due to the limitations of manual transcription and phonetic labelling, which are time consuming, costly and prone to error if involved thousands of speech files; thus, researchers prefer automated approach [13-17].

The use of automatic transcription and phonetic labelling in transcribing and labelling speech is now pervasive as the considerable gains in time and cost of automatic method made it an alternative way to handle limitation of manual transcription [13, 18, 19]. This alternative approach can be performed faster compared to manual transcription [18, 20, 21, 22].

II. Dyslexia And Highly Phonetically Similar Errors

Dyslexia is caused by deficits in the phonological parts in the brain that it hinders the development of literacy skills [37, 38] (and sometimes affects other skills too, for example, writing, mathematics or motor skills). There has been recent evidence that shows dyslexics suffer a delay in developing into pre-literacy and emergent literacy stages of a child's life [39]. The prevalence of dyslexia has inspired this work to be taken, considering the potential that dyslexic children can do well in academics and literacy learning as their cognitive ability is at par with their normal peers. Nevertheless, these children have shown little or slow improvement after conventional intervention [40]. Thus, it is viewed that speech-enabled technology, such as an automatic reading tutor or interactive reading tutor could facilitate them to learn to read better. In order to do so, automatic speech recognition is important to provide immediate intervention during reading [44]. Hence, automatic transcription and labelling are deemed as important too. The challenge here lies in the highly phonetically similar errors made when they are reading that hinders ASR to produce high accuracy.

Phonetically similar errors, in this case, refers to the reading mistakes often made by the children when they are reading (refer to Table 1). For example, the word 'pada' is often read as 'bapa' or vice versa due to the lookalike feature of the word (mirror letters of b and p) and the phonetic similarity of the letters. These errors are difficult to be recognized as they are very similar in sound and in appearance in the spectrogram, thus affecting the ability to transcribe, segment, and phonetically label them.

III. Automatic Transcription And Phonetic Labelling

Transcription and phonetic labelling involve transforming speech into small units called phonetic symbols. In this study, Worldbet is used as it covers world languages [23]. Each approach produced 585 phoneme files, i.e. the segmentation and phonetic labelling files, of 585 dyslexic children's read speech in Malay. The 585 speech files are selected randomly from existing corpus [24].

A. Manual Transcription and Labelling

Manual transcription refers to the process whereby speech files are transcribed and phonetically labelled manually. Fig. 1 illustrates the spectrogram of a speech sample with its segmentation and phonetic labels. To perform manual transcription, a few steps need to be performed: 1) the recorded speech file is opened to view its spectrogram; 2) the transcriber needs to listen to the speech file and manually segment the spectrogram and then label it according to its phonetic representation; 3) repeat step two a few times until satisfied to ensure accurate segmentation and labelling has been performed; 4) produce the transcript based on the labelling. Fig. 1 illustrates an example of a speech spectrogram for the word 'cantik'. Based on the spectrogram, transcriber listens and decides where each phoneme should be segmented and what phoneme belongs to that segment. In Fig. 1, the segments are denoted by the bottom row, where each segment is labelled with its corresponding phonetic symbol, representing the suitable phoneme. Note that phoneme is the smallest unit in a language, thus it can be extremely confusing when one has to decide if the sound is very similar, e.g. `bh` and `ph`. Imagine if we have to transcribe a large amount of speech files for ASR, the task shall be overwhelming and thus error-prone.

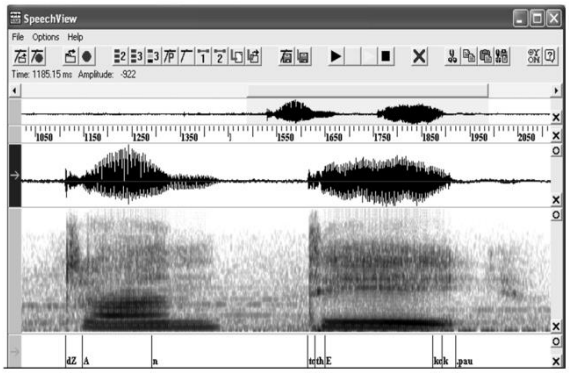

Figure 1: A sample of speech file of the word 'cantik'.

In this study, manual transcriptions act as a benchmark for examining the acceptable accuracy of automatic transcription and phonetic labelling. This is because researchers believed that manual transcription method are more accurate due to the use of human transcribers that ensures that transcription and phonetic label are perceptually valid [8]. Furthermore, manual transcription requires human transcribers to hear each individual phoneme of a word prior to performing segmentation and phonetic labelling.

B. Automatic Transcription and Labelling

To generate automatic transcription and phonetic labelling, force alignment is performed where existing ASR engine is used to force align the new 585 speech files. Forced alignment is an approach to perform automatic transcription and phonetic labelling based on existing lexical model [24]. Many speech recognition systems have used the technique of Viterbi alignment algorithm or the forced alignment [35, 36]. These systems have the ability to recognize pronunciation variation or multiple pronunciations of a spoken word. The output of this process is a total of 585 phoneme files in .phn format. These files are important input files prior to training an ASR engine to build one that could potentially 'listen' to the highly phonetically similar errors often made by dyslexic children when they read aloud some isolated words in Malay. As aforementioned, the errors made when reading is indeed a challenge for ASR to perform good recognition. Fig. 2 shows the output of automatic transcription and phonetic labelling for the word 'cantik' (beautiful).

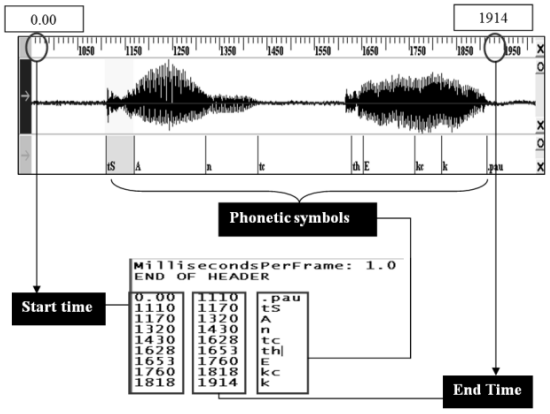

Figure 2: The word 'cantik' with its segments and phonetic labels generated automatically.

Every .phn files contain segmentations of the speech that is labelled with phonetic labels, as well as the starting time and ending time for each phoneme. There are three columns for each .phn file where the first column represents the start time in milliseconds (ms), the second column represents the end time in milliseconds (ms), and the third column represents the phonetic symbols for that segment. The first two lines of the file are headers which define the length of a "frame" in milliseconds (ms). The rest of the files consist of two numbers that define a frame range, and a label that applies to that region.

Obviously performing automatic transcription and segmentation and labelling saves a lot of time and effort when compared with the manual approach. However, we need to examine and compare the performance of automatic transcription and labelling when dealing with such highly phonetically similar errors in terms of its accuracy. Prior to using the automatically generated input for training an ASR engine, we first measure its accuracy by comparing the input files generated to the one which we manually transcribe and segment and label. For this purpose, the Levenshtein Distance algorithm is employed to measure the distance between the start point and the end point of the automatically generated phoneme files against the manual counterparts. The results obtained are promising with 95% similarity in terms of the phonetic labels and 65% similarity in terms of the duration of phonetic segmentation.

C. Training ASR Engine

To measure the accuracy of the transcription and phonetic labelling, using the manual approach as a benchmark, two ASR engines were trained using HMM-ANN as the hybrid method gives better accuracy [28-32]. In this training, the lexical model is improved to cater for the input of the training, i.e. the 585 speech files with the corresponding transcription and labelling generated manually as well as automatically. Thus, there are two ASR engines trained; one by training input files produced manually and another by training input files generated automatically. The training involved 30 networks iterations for both transcriptions files. The process iterates until optimum accuracy is achieved on the development dataset and only then it is tested on the testing dataset to evaluate final network. The final network with the highest recognition accuracy on test dataset is regarded as the optimum engine, the one that can be used for further evaluation of WER and FAR.

IV. Results and Discussion

The accuracy of ASR engines was measured using the standard metrics i.e. recognition rate, WER and FAR. In this section, the discussions emphasize on the results of training from two ASR engines: one that is built by training the automatically generated phoneme files and transcription files, the other is built by training the manually generated files.

A. Training Results

After series of training, the recognition accuracy on test dataset is obtained for both ASR engines. The training is stopped when the accuracy percentage starts to decrease. Table 2 depicts the results of training using input files from both manual and automatic transcription and labelling approach.

Table 2
Series of Trainings Performed on Manually and Automatically Generated Input Files

	Manual		Automatic	
	Best %	Best network	Best %	Best network
1	54.33	Wordsnet.25	52.34	Wordsnet.12
2	58.27	Wordsfa2net.30	56.25	Wordsfa2net.4
3	62.20	Wordsfa3net.29	61.72	Wordsfa3net.1
4	61.42	Wordsfa4net.18	59.38	Wordsfa4net.24
5	**76.29**	**Wordsfa4net.29**	**76.04**	**Wordsfa3net.9**

Referring to the results presented in Table 2 for manually generated input training, the accuracy from the first until the third training showed improvements which give 54.33% and then increased to 58.27%. Subsequently, the result of training using manually generated input files increased 3.93% in the third training given 62.20%. However, the performance of fourth training slightly decreased to 61.42%. Thus, the training on development dataset for manual transcription and labelling is stopped. The fifth training is the final results for ASR engine using test dataset. The training in the test dataset used Wordsfa3net.29 from the third training as the input network for ASR engine using manual transcription and labelling. The result of ASR engine trained on manually generated transcription and segmentation and labelling is 76.29%.

A series of training was also conducted using automatic transcription and phonetic labelling. Based on Table 2, the results of ASR engine using automatic transcription and phonetic labelling is at par with the results of training on manually generated transcription and labelling. The first until the third training results showed enhancement of 52.34%, 56.25% and 61.72% respectively. However, in the fourth training, the result decreased to 59.38%. Therefore, the best results of training on development dataset for automatic transcription and phonetic labelling is also given by the third training. Thus, we used the third network, i.e. Wordsfa3net.1 as the network to train the test dataset. The accuracy of ASR engine for automatic transcription and phonetic labelling is 76.04% which is similar to that of ASR using manual transcription and labelling.

B. Discussion

From the results, it is observed that the performance of ASR engine that is trained on automatically generated input files is very much similar to its manual counterpart. The results for both ASR engines are 23.9% WER for automatically generated input and 23.7% WER for the

manually generated input. Thus, we can conclude that the ASR engine trained on automatically generated input files, i.e. the transcription and phonetic labels, performs on par with the manual one, which we regard as the benchmark for evaluation. The FAR for both ASR engines gives slightly the same percentage i.e. 17.9% for the manual one and 18.1% for the automatic one given only 0.2%. Since FAR is defined as a number of correct reading recognized as incorrect over the total number of correct readings in percentile, it showed no significant reduction if only 0.02% WER is observed in the comparison of both engines.

Given the observation above, it is shown that both ASR engines have similar WER and thus automatic transcription and phonetic labelling can potentially be used to transcribe and phonetically label the dyslexic children's read speech towards the development of an ART or IRT. In the WER evaluation, the recognition accuracy performances depend on the ability to recognize the words. Hence, the lower the percentage is the better. Therefore, the lowest WER for automatic transcription and phonetic labelling is 23.9% and manual transcription with 23.7%. The WER is influenced by highly phonetically similar errors from dyslexic children's reading, which is why the WER is somewhat higher when compared with other researchers that deal with normal speech such as in [33, 34]. Thus, phonetically similar errors in dyslexic children's read speech affected not only the recognition accuracy of WER and FAR, but also the performance of automatic transcription and labelling just as it affected manual transcription and labelling. However, the automatic approach can still be used to automate and replace the tedious, error-prone process of the manual approach towards the development of ASR for the purpose of ART or IRT.

Although the WER and FAR are somewhat higher when compared to the performance involving normal speech, for example in [41], [42] or [43], when it comes to highly phonetically similar errors produced by dyslexic children reading, we can say that this is good enough to facilitate automatic transcription and segmentation and labelling process prior to building an ASR for the purpose of ART or IRT. The small difference in terms of percentage (0.02% of WER) between manual and automated transcription and labelling suggest that automated version can, therefore, be used as alternative towards this effort. As current performances when it comes to children's speech remains a challenge [42], we believe that this is a start towards enabling an interactive reading tool to facilitate dyslexic children reading in Malay so that they could learn to read better and be more engaged in reading activities. We also believed that ASR technology that is going to be developed for them should also be able to 'listen' to their reading by accepting or tolerating with the spoken Malay influence in their reading. In another study [45], we have found that the word with less variability in pronunciation, i.e. not really influenced spoken Malay, scored the highest segmentation similarity with 79% accuracy. This suggests that, in order for an ASR to 'listen' to dyslexic children's reading with highly phonetically similar errors, the engine should adapt to the spoken Malay especially words that have higher variability in pronunciation, such as "betul", "kampung", and "umur". In future, the spoken Malay will be modeled into the lexical model allowing the computer to 'listen' better by adapting to the variability of the pronunciation of a word and potentially improve automatic segmentation of the highly phonetically similar speech data.

V. Conclusion

The study was set out to investigate the accuracy of ASR engine when using automatic transcription and phonetic labelling of dyslexic children's read speech in Malay. The challenge lies in the nature of the children's readings that normally contain highly phonetically similar errors. As manual transcription and labelling are often regarded as the best, it serves as a benchmark in evaluating whether automatic transcription and segmentation and labelling is a potential approach to replace the tedious, time consuming manual approach. The accuracy of ASR engine using automatic transcription and phonetic labelling has been evaluated to see if it is acceptable for the development of ASR engine specifically for dyslexic children's reading in Malay. With the results of only 0.02% difference between the manual and automatically generated transcription and labelling, we can conclude that even though the readings contain highly phonetically similar errors, their effects on transcription and labelling is slightly the same, be it for manual or automatic approach. Hence, the ASR trained on manually generated or automatically generated transcription and labelling did not significantly differ. Thus, in dealing with highly phonetically similar errors in dyslexic children's readings, the WER and FAR are not as low as normal speech but it does indicate that automatic transcription and labelling can be used to train and develop an ASR to handle dyslexic children's reading.

Acknowledgment

The authors would like to thank headmasters and dyslexia teachers from Sekolah Kebangsaan Jalan Datuk Kumbar, Alor Setar, Kedah and Sekolah Kebangsaan Taman Tun Dr. Ismail (2), Kuala Lumpur and the students who are involved, directly and indirectly, in this research.

References

[1] T. Athanaselis, S. Bakamidis, I. Dologlou, E. N. Argyriou, A. Symvonis, "Making assistive reading tools user friendly: A new platform for Greek dyslexic students empower by automatic speech recognition". *Multimedia Tools and Application*, vol. 68, no. 3, pp. 681-699, 2014.

[2] M. Taileb, R. Al-Saggaf, A. Al-Ghamdi, M. Al-Zebaidi, S. Al-Sahafi, "YUSR: Speech recognition software for dyslexics". *Design, User Experience, and Usability, Health, Learning, Playing, Cultural, and Cross-Cultural User Experience*, vol. 8013, pp. 296-303, 2013.

[3] J. S. Pedersen, L. B. Larsen, "A Speech Corpus for Dyslexic Reading Training," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC), European Language Resources Association*, pp. 2820-2823, 2010.

[4] H. Husniza, J. Zulikha, "Dyslexic children's reading pattern as input for ASR: Data, analysis, and pronunciation model," *Journal of Information and Communication Technology*, vol. 8, pp. 1-13, 2009.

[5] X. Li, L. Deng, Y. C. Ju, A. Acero, " Automatic children's reading tutor on hand-held devices," in *Proceedings of Annual Conference of the International Speech Communication Association*, vol. 9, pp. 1733-1736, 2008.

[6] C. Cucchiarini, H. Strik, "Automatic phonetic transcription: An overview," in *Proceedings of the International Congress of Phonetic Sciences (ICPhS), Barcelona*, vol. 15, pp. 347–350, 2003.

[7] J. P. Goldman, "EasyAlign: An automatic phonetic alignment tool under Praat," in *Proceedings of Annual Conference of the International Speech Communication Association, Folorence*, vol. 12, pp. 3233-3236, 2011.

[8] A. Dupuis, "Automatic transcription of audio files and why manual transcription may be better," Retrieved March 23, 2015, from:

http://www.researchware.com/company/blog/368-automatictranscription.html. 2011.

[9] K. Yu, M. Gales, L. Wang, P. C. Woodland, "Unsupervised training and directed manual transcription for LVCSR," *Speech Communication*, vol. 52, no. 7, pp. 652-663, 2010.

[10] M. Dinarelli, A. Moschitti, G. Riccardi, "Concept Segmentation and Labeling for Conversational Speech," in *Proceedings of Annual Conference of the International Speech Communication Association*, vol. 10, pp. 2747-2750, 2009.

[11] T. J. Hazen, "Automatic alignment and error correction of human generated transcripts for long speech recordings," in *Proceedings of International Conference on Spoken Language Processing, Pittsburgh*, vol. 9, pp. 1606-1609, 2006.

[12] T. Bauer, L. Hitzenberger, L., Hennecle, "Effects of manual phonetic transcriptions on recognition accuracy of streetnames," in *Proceedings of the International Symposiums for Information Swissenschaft (ISI)*, vol. 8, pp. 21-25, 2002.

[13] J. Yuan, N. Ryant, M. Liberman, A. Stolcke, V. Mitra, W. Wang, "Automatic phonetic segmentation using boundary models," in *Proceedings of Interspeech Annual Conference of the International Speech Communication Association*, pp. 2306-2310, 2013.

[14] H. Husniza, Y. Yuhanis, K. Siti Sakira, "Speech Malay language influence on automatic transcription and segmentation," in *Proceedings of the International Conferences on Computing and Informatics, ICOCI, Sarawak, Malaysia*, vol. 4, pp. 132-137, 2013.

[15] B. Schuppler, M. Ernestus, O. Scharenborg, L. Boves, "Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions," *Journal of Phonetics*, vol. 39, no. 1, pp. 96-109, 2011.

[16] C. Van Bael, L. Boves, H. Heuvel, H., Strik, "Automatic Phonetic Transcription of Large Speech Corpora," *Computer Speech & Language*, vol. 21, no. 4, pp. 652-668, 2007.

[17] J. P. Hosom, "A Comparison of speech recognizers created using manually-aligned and automatically-aligned training data," Technical Report CSE-00-02, Oregon Graduate Institute of Science and Technology, Center for spoken Language Understanding, Beaverton, 2002.

[18] F. Cangemi, F., Cutugno, B. Ludusan, D. Seppi, C. D. Van, "Automatic speech segmentation for Italian (ASSI): Tools, models, evaluation, and applications," in *Proceedings of the Associazione Italiana di Scienze della Voce (AISV), Lecce, Italy*, vol. 7, pp. 337-344, 2011.

[19] E. A. Kaur, E. T. Singh, "Segmentation of continuous Punjabi speech signal into syllables," in *Proceedings of the World Congress on Engineering and Computer Science*, vol. 1, pp. 20-22, 2010.

[20] V. Silber, N. Geri, "Can automatic speech recognition be satisfying for audio/video search? Keyword-focused analysis of Hebrew automatic and manual transcription," *Online Journal of Applied Knowledge Management*, vol. 2, no. 1, pp. 104-121, 2014.

[21] M. Sperber, "Efficient speech transcription through respeaking," Master's Thesis, Karlsruhe Institute of Technology Department of Computer Science, 2012.

[22] J. D. Williams, I. D. Melamed, T. Alonso, B. Hollister, J. Wilpon, "Crowd-sourcing for difficult transcription of speech," in *Proceedings of IEEE Workshop, Automatic Speech Recognition and Understanding (ASRU)*, pp. 535-540, 2011.

[23] L. J. Hieronymus, "ASCII Phonetic Symbols for the world's Languages: Worldbet," Bell Laboratories manuscript, 1993.

[24] H. Husniza, "Automatic speech recognition model for dyslexic children reading in Bahasa Melayu," Doctoral dissertation, Universiti Utara Malaysia, 2010.

[25] M. Dinarelli, A. Moschitti, G. Riccardi, "Concept Segmentation and Labeling for Conversational Speech," in *Proceedings of Annual Conference of the International Speech Communication Association*, vol. 10, pp. 2747-2750, 2009.

[26] T. J. Hazen, "Automatic alignment and error correction of human generated transcripts for long speech recordings," in *Proceedings of International Conference on Spoken Language Processing, Pittsburgh*, vol. 9, pp. 1606-1609, 2006.

[27] D. Gibbon, "Part 1: Spoken language system and corpus design," in *Handbook of standards and resources for spoken language systems*, Berlin: Mouton de Gruyter, 1997.

[28] M. Frikha, A. B. Hamida, "A comparative survey of ANN and hybrid HMM/ANN architectures for robust speech recognition," *American Journal of Intelligent Systems*, vol. 2, no. 1, pp. 1-8, 2012.

[29] H. F. Ong, A. M. Ahmad, "Malay Language Speech Recognizer with Hybrid Hidden Markov Model and Artificial Neural Network (HMM/ANN)," *International Journal of Information and Education Technology*, vol. 1, no. 2, pp. 114-119, 2011.

[30] H. Husniza, J. Zulikha, "Dyslexic children's reading pattern as input for ASR: Data, analysis, and pronunciation model," *Journal of Information and Communication Technology*, vol. 8, pp. 1-13, 2009.

[31] R. Fadhilah, R. N. Ainon, "Isolated Malay speech recognition using Hidden Markov models," in *Proceedings of the International Conferences on Computer and Communication Engineering*, pp. 721-725, 2008.

[32] H. A. Bourlard, N. Morgan, "Connectionist speech recognition: A hybrid approach," *Springer Science & Business Media*, vol. 247, 2012.

[33] J. P. Hosom, "A Comparison of speech recognizers created using manually-aligned and automatically-aligned training data," Technical Report CSE-00-02, Oregon Graduate Institute of Science and Technology, Center for spoken Language Understanding, Beaverton, 2002.

[34] H. Sarma, N. Saharia, U. Sharma, "Development of Assamese speech corpus and automatic transcription using HTK," *Advances in Signal Processing and Intelligent Recognition Systems*, vol. 264, pp. 119-132, 2014.

[35] F. Schiel, "Automatic phonetic transcription of non-prompted speech," in Proceedings of International Congress on Phonetic Science, pp. 607-610, 1999.

[36] S. Rapp, "Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov Models: An Aligner for German," 1995.

[37] M. Melby-Lervåg, S.-A.H. Lyster, C. Hulme, "Phonological skills and their role in learning to read: a meta-analytic review," *Psychology Bulletin*, vol. 138, pp. 322-352, 2012.

[38] F.R. Vellutino, J.M. Fletcher, M.J. Snowling, D.M. Scanlon, "Specific reading disability (dyslexia): what have we learned in the past four decades?" *Journal of Child Psychology & Psychiatry*, vol. 45, pp. 2–40, 2004.

[39] F. Morkena, T. Hellanda, K. Hugdahl, K. Spechta, "Reading in dyslexia across literacy development: A longitudinal study of effective connectivity," *Neuroimage*, vol. 144, pp. 92-100, 2017.

[40] L. Caroline, L. Cupple, "Thinking outside the boxes: Using current reading models to assess and treat developmental surface dyslexia," *Journal of Neuropsychological Rehabilitation*, vol. 27, pp. 149-195, 2017.

[41] T. Baumann, C. Kennington, J. Hough, D. Schlangen, "Recognising conversational speech: What an incremental ASR should do for a dialogue system and how to get there," Lecture Notes in Electrical Engineering, vol. 999, pp 421-432, 2016.

[42] J. Kennedy, S. Lemaignan, C. Montassier, P. Lavalade, B. Irfan, F. Papadopoulos, E. Senft, T. Belpaeme, "Child Speech Recognition in Human-Robot Interaction: Evaluations and Recommendations," 2016.

[43] T. Athanaselis, S. Bakamidis, I. Dologlou, E. N. Argyriou, A. Symvonis, "Making assistive reading tools user friendly: A new platform for Greek dyslexic students empowered by automatic speech recognition," *Multimedia Tools and Applications*, vol. 68, issue 3, pp. 681-699, 2014.

[44] H. Husni, Z. Jamaludin, "ASR technology for children with dyslexia: Enabling immediate intervention to support reading in Bahasa Melayu," *US-China Education Review*, vol. 6, no. 6, pp. 64-70, 2009.

[45] H. Husni, Y. Yusof, S. S. Kamaruddin, , "Spoken Malay language influence on automatic transcription and segmentation," in *Proceedings of the International Conference on Computing and Informatics, ICOCI*, 2013.